基于联合测序分析技术挖掘红苞凤梨 lncRNAs 信息

胡 豪[1]，蔺 珍[1]，薛彦斌[1]，毛美琴[1]，向奕璇[1]，刘加文[1]，周徐子鑫[1]，马 均[1,2*]

（1，四川农业大学 风景园林学院，成都 611130；2，四川农业大学

园林研究所，成都 611130）

摘要：为了揭示 lncRNA 在红苞凤梨嵌合叶片形成和生长发育过程中的调控作用机制，以金边红苞凤梨为材料，采用 Iso-seq 2005 测序和 SMRT 三代全长转录组测序联合测序分析技术，分析挖掘红苞凤梨 lncRNA 信息。结果表明，鉴定得到 6,018 条 lncRNA，包含了 3,298 个基因间 lncRNAs，717 个反义 lncRNAs，890 个内含子 lncRNAs 和 1,109 个正义 lncRNAs，数据量较二代测序有了极大的提高。结构分析表明，红苞凤梨 lncRNA 的总体表达丰度低于 mRNA；序列长度在 400~1200 nt 区间比例高于 mRNA，而在>1600 nt 区间，lncRNA 分布的比例显著小于 mRNA；lncRNA 中的外显子数量总体少于 mRNA，开放阅读框长度总体上也短于 mRNA。差异表达分析表明，在全绿、全白叶片发育过程中鉴定到 1710 个差异表达 lncRNA。靶基因预测结果表明，5441 个 lncRNAs 通过 Cis 作用方式预测到靶基因，1544 个 lncRNAs 通过 Trans 方式预测到靶基因。靶基因的功能注释和富集分析显示，差异表达 lncRNA 的靶基因主要作为酶蛋白参与调节叶片代谢活动和信号转导等方面，与叶片的颜色形成、光合作用和生长发育密切相关。该文鉴定出的 lncRNA 信息以及对其结构和功能的分析，为红苞凤梨以及凤梨科其它植物的 lncRNA 表观遗传调控机理研究提供了数据基础，筛选出的差异表达 lncRNA 在金边红苞凤梨叶片嵌合性状的形成和生长发育中具有重要的调控作用。

关键词：红苞凤梨，Iso-seq 测序，SMRT 全长转录组测序，lncRNA 鉴定

lncRNAs identification of *Ananas comosus* var. *bracteatus* by hybrid sequencing

---

收稿日期：2019-12-27

基金项目：国家自然科学基金（3157030554，31770743）[Supported by the National Nature Science Foundation of China（3157030554，31770743）]。

作者简介：胡豪(1996-)，男，重庆，四川农业大学，硕士生，园林植物培育与遗传育种，1095142644@qq.com。

***通信作者：**马均，博士，教授，研究方向为园林植物培育与遗传育种，junma365@hotmail.com。

HU Hao[1], LIN Zhen[1], XUE Yanbin[1], MAO Meiqin[1], XIANG Yixuan[1], LIU Jiawen[1], ZHOUXU Zixin[1], MA Jun[1,2*]

（1. *College Of Landscape Architecture*，*Sichuan Agricutural University*，Chengdu 611130，China；2. *Academy Of Agriculture*，*Sichuan Agricutural University*，Chengdu 611130，China）

Abstract: In order to reveal the regulation function of lncRNAs on the chimeric character formation and development of the leaf of Ananas comosus var. Bracteatus, Iso-seq 2005 sequencing and SMRT full-length transcriptomesequencing were applyed to identify lncRNAs of Ananas comosus var. bracteatus. In this study, 6 018 lncRNAs were identified, containing 3 298 intergenic IncRNAs, 717 antisense IncRNAs, 890 intron IncRNAs and 1 109 sense IncRNAs, which were greatly improved compared with Iso-seq 2005 sequencing. Structural analysis showed that the overall expression level of lncRNA was lower than that of mRNA. The transcript length distribution of lncRNA in the range of 400-1200 nt was higher than that of mRNA, while in the > 1 600 nt, the proportion of lncRNA distribution was significantly lower than that of mRNA. The number of exons in lncRNA was generally less than that of mRNA, and the open reading frame was also shorter in length than that of mRNA. For analysis of differential expression, 1 710 differentially expressed lncRNAs were identified during the development of complete green and complete white leaves. Target gene prediction results showed that 5 441 lncRNAs were predicted target genes by Cis action, and 1 544 lncRNAs were predicted target genes by Trans action. Functional annotation and enrichment analysis of target genes revealed that the target genes of differentially expressed lncRNAs mainly act on metabolic activities and signal transduction of leaves as enzyme proteins, and were closely related to leaf color formation, photosynthesis and leaf growth. The lncRNA information identified in this paper and, as well as the analysis of its structure and functions , provide a data basis for the study of the epigenetic regulation mechanism of lncRNA in Ananas comosus var. bracteatus and other plants in Bromelia. The identified differentially expressed lncRNAs play important roles in the chimeric character formation and development of leaf of *Ananas comosus* var. bracteatus.

Key words: *Ananas comosus* var. *Bracteatus,* Iso-seq, SMRT full length transcriptome sequence, lncRNA identification

红苞凤梨（*Ananas comosus* var. *bracteatus*）因其叶片绿白镶嵌、花果颜色艳丽且观赏期长，已成为一种重要的新型观赏植物。红苞凤梨自交不亲和，在生产中以吸芽进行繁殖，繁殖系数低、苗木整齐度差，限制了红苞凤梨的规模化应用。组织培养能快速繁殖红苞凤梨，但繁殖过程中叶片嵌合性状不稳定，再生植株叶片常失去嵌合性状而变为全绿植株（曹莉，2011）。细胞白化突变是金边嵌合叶片形成的基础，研究红苞凤梨叶片细胞白化突变的分子机理，对于揭示红苞凤梨嵌合性状的形成机理，提高嵌合性状的稳定性，培育新的嵌合性状品种具有重要的理论和实践意义。

本课题组前期研究结果表明，红苞凤梨白化细胞中叶绿素含量极显著下降，但叶绿素合成代谢的结构基因表达上调（Li et al.，2017；Xue et al.，2019），说明转录后调控在红苞凤梨细胞白化失绿、金边嵌合性状形成中发挥了重要调控作用。LncRNA 具有类似 mRNA 的结构特征，可在多个层面调控目标基因的表达。可作为信号分子、诱饵分子、引导分子以及支架分子，在表观遗传、转录调控、转录后调控等多个水平发挥功能（Zhang et al.，2018）。目前，lncRNA 在人类和动物中研究较为广泛，与人类疾病的发生及生物体的生长发育密切相关(余铖亮等，2015；JOHNSON，2012；王艳芳等，2018)。而植物 lncRNA 的研究还处于起步阶段。研究结果表明 lncRNA 在植物的开花诱导（Csorba et al.，2014）、花粉发育（DING et al.，2012）、逆境胁迫（Qin et al.，2017）中具有重要功能，然而其具体的作用机制及调节功能等尚不清楚。由于红苞凤梨没有基因组数据，且 Iso-seq 二代测序技术读长较短，使得红苞凤梨非编码 RNA 的挖掘具有一定的局限性。随着测序技术的发展，SMRT(Single-molecule real-time)三代测序技术的出现，因其无需进行 PCR 扩增，大幅降低了因 PCR 反应引入的碱基错误，操作更为简单等优点，已得到广泛应用（Flusberg et al.，2010）。目前，SMRT 三代测序技术在基因组、甲基化识别、SNP 的鉴定、基因重测序和转录组学等方面的优势越来越明显（Smith et al.，2012；Guo et al.，2018；）。此外，SMRT 测序技术得到几千 KB 的数据，读长显著增长，大大减少了测序后的 Contig 数量，使得基因组和转录组的组装得到极大改善（English et al.，2012）。但是第三代测序技术也存在一定的缺点，测序错误率普遍偏高，测序产生的错误率可高达 15%（Koren et al.，2017）。所以，采取二代测序和三代测序联合分析已成为当下基因组研究的主要方法。一般有两种联合手段，可以选择以三代测序为主，进行组装，再通过二代测序得到的高质量短片段对三代的数据进行碱基纠错和矫正；也可以选择以二代测序为主，用三代测序得到的长片段 reads 进行辅助组装（马建超，2018）。最为通常的手段是利用短的但准确度高的二代数据去辅助校正长的但准确度较低的三代数据，并在此基础上进行混合组装。经验证，通过这种混合组装

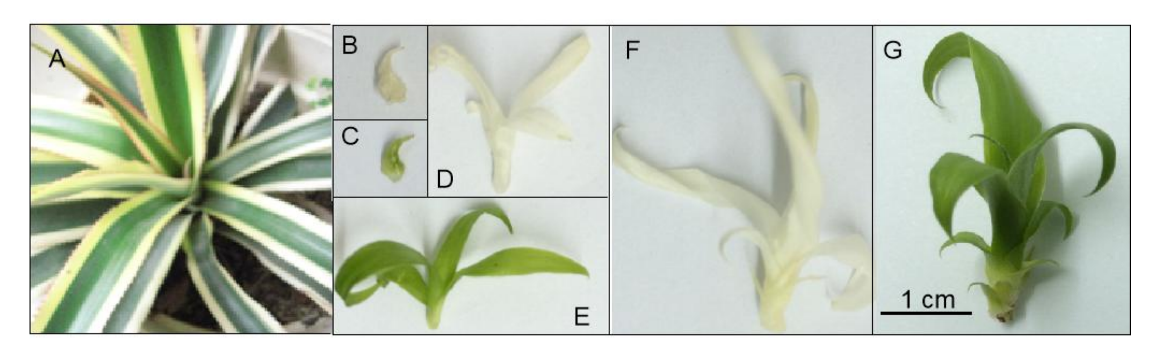
校正方法得到的数据准确性可达 99%（马东娜等，2018）。因此，这种“2 + 3”的联合模式已被广泛认可并应用于动物、植物和微生物的基因组研究（Koren et al.，2012；Hackl et al.，2014；徐伟南等，2018）。

本研究以红苞凤梨为材料，利用二代和三代测序数据混合组装校正方法，鉴定红苞凤梨叶片中存在的 lncRNA，分析在全绿和全白突变叶片生长发育过程中差异表达的 lncRNA，通过靶基因的功能注释和富集分析，揭示 lncRNA 在红苞凤梨叶片细胞失绿白化及生长发育过程中的作用。研究结果为红苞凤梨 lncRNAs 表观遗传调控机理研究提供数据基础。

1. 材料与方法

1.1 试验材料

本研究以金边红苞凤梨茎段为外植体，通过组织培养获得的全白和全绿植株为试验材料。选取长势均一的未展叶期、4~5 叶期和 10~12 叶期 3 个发育阶段的全绿及全白植株各 10 株的叶片作为样本提取 RNA（图 1）。样本取下后立即转入液氮速冻，后储存于-80℃用于提取 RNA 进行 Iso-seqRNA2500 二代转录组测序。



注：**A**.红苞凤梨野生型嵌合植株；**B**.未展叶的全白芽（CW1）；**C**.未展叶的全绿芽（CG1）；**D**.全白苗具 4~5 片叶（CW2）；**E**.全绿苗具 4~5 片叶（CG2）；**F**.全白苗具 10~12 片叶（CW3）；**G**.全白苗具 10~12 片（CG3）。标尺=1 cm。

Note: **A**. *Ananas comosus* var. *Bracteatus* wild-type chimeric plants; **B**. Un-leaved complete white buds (CW1); **C**. Un-leaved complete green buds (CG1); **D**. Complete white plantlets with 4~5 Leaf (CW2); **E**. Complete green plantlets with 4~5 leaves (CG2); **F**. Complete white plantlets with 10~12 leaves (CW3); **G**. Complete white plantlets with 10~12 pieces (CG3). Ruler=1 cm.

图 1 用于测序的红苞凤梨全绿苗和全白苗（Xiong et al.，2018）

Fig. 1 Different development stages of green and albino shoots of *Ananas comosus* var. *bracteatus*

1.2 红苞凤梨 RNA 提取及二代 Illumin Hiseq2500 测序文库构建

1.2.1 红苞凤梨 RNA 提取及检测

取-80℃冻存的红苞凤梨样品，采用 LABGENETM plant RNA Isolation kit 多糖多酚植物 RNA 提取试剂盒分离总 RNA，操作方法参考使用说明。高质量的 RNA 是实验成功的基础，为保证测序的准确性，对样品 RNA 进行了检测，分别采用荧光定量仪（Qubit 2.0）、微量分光光度计（Nanodrop）、生物分析仪（Aglient 2100）、电泳方法检测 RNA 样品的纯度、浓度、完整性和是否有基因组 DNA 污染等，达到质控要求的 RNA 样品用于后续试验。

1.2.2 Illumina Hiseq2500 cDNA 文库构建及上机测序

利用 epicentre Ribo-ZeroTM 试剂盒去除样品中的 rRNA。以 rRNA-depleted RNA 为模板，用六碱基随机引物（random hexamers）合成 cDNA 第一条链和第二条链。cDNA 纯化后进行末端修复、加 A 尾并连接测序接头，然后用 AMPure XP beads 进行片段大小选择。降解含 U 链，通过 PCR 富集得到 cDNA 文库。文库构建后，经 Qubit2.0 定量， Agilent 2100 检测文库的 insert size 质量；采用 QPCR 对文库的浓度（文库有效浓度＞2nM）进行准确定量，完成库检。库捡完成后，采用 Illumina Hiseq2500 平台对文库进行测序。

1.3 数据质量监控

测序所得 reads 的过滤与修剪是保证分析数据可靠性的关键，结合前期 SMRT 测序结果（MA et al.，2018）进行混合组装校正后，删除 raw reads 中包含 adapter、ploy-N 的 reads 和低质量的 reads 以获得 clean reads。以菠萝基因组（Acomosus_321_v3https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Acomosus_er）为参考基因组，利用 TopHat v2.0.9（KIM et al.，2013）软件对 clean reads 进行比对。采用 Scripture（beta2）（LANGMEAD et al.，2009）和 Cuffiinks（v2.1.1）软件对每个样品比对上的 reads 进行装配获得转录本。

1.4 转录本表达水平和编码潜能分析

使用 Cufflinks 软件的 Cuffdiff 组件，对转录本表达水平进行分析。根据 lncRNA 的编码的特点进行基本筛选：选择长度≥200bp，Exon 个数≥2 及 FPKM≥0.1 的转录本。

因 lncRNA 不编码蛋白，因此，通过对基本筛选得到的转录本进行编码潜能筛选，判断其是否具有编码潜能，从而可以判定该转录本是否为 lncRNA。主要利用 CPC 分析、CNCI 分析、CPAT 分析、pfam 蛋白结构域四种分析方法分析 lncRNA 的编码能力。去掉基本筛选

中具有潜在编码能力的转录本，余下的即为预测的 lncRNA。

通过与已知 mRNA 进行比较，利用 cuffcompare 分析结果中的 class codes 对筛选的 lncRNA 进行分类。

1.5 lncRNA 靶基因预测及功能富集分析

基于 lncRNA 与其靶基因的作用方式（cis 和 trans），采用两种预测方法：第一种是根据 lncRNA 与 mRNA 的位置关系预测 lncRNA 的靶基因，即 lncRNA 100kb 范围内的邻近基因为其靶基因。第二种是使用 LncTar 靶基因预测工具进行预测，其原理是利用 lncRNA 与 mRNA 之间碱基互补配对产生的作用来预测。

对差异表达 IncRNA 的靶基因利用 KEGG(Kyoto Encyclopedia of Genesand Genomes)、GO(Gene Ontology)、NR (NCBI non-redundant protein sequences)、COG (Clusters of Orthologous Groups of proteins)和 Swiss-Prot 数据库进行功能注释和富集分析，分析结果显著性用 p 值表示。

1.6 差异表达基因分析

以差异倍数（Fold Change）≥2 且错误发现率 FDR（False Discovery Rate）<0.05 作为差异表达的筛选标准，使用 EBseq 分析 6 个样本间差异表达的 lncRNAs 及 mRNA。

2. 结果分析

2.1 测序数据与参考基因组比对效率分析

长链非编码数据利用率的直接体现就是比对效率，即 Mapped Reads 占 Clean Reads 的百分比。经 SMRT 全长转录组数据（NCBI 提交号 PRJNA564223）（MA et al.，2018）修正后，6 个样品的 Reads 与所选参考基因组的比对效率为 67.74%~78.58%，比单独使用 Iso-seq 二代测序数据的比对效率提高了约 5%（蔺珍，2019）（表 1）。说明三代数据的修正，有效提高了 lncRNA 测序数据的利用效率，有利于进一步深入挖掘红苞凤梨 lncRNA 信息。

表 1　LncRNA 测序数据与参考基因组比对结果统计表

Table 1 Statistical table of results comparing LncRNA sequencing data with reference genomes

BMK-ID	Total Reads	Mapped Reads	Uniq Mapped Reads	Multiple Mapped Reads	Reads Map to '+'	Reads Map to '-'

CG1	79076522	55189580(69.79%)	49512630(62.61%)	5676950(7.18%)	26221992(33.16%)	26114494(33.02%)
CG2	58211966	43212038(74.23%)	38969713(66.94%)	4242325(7.29%)	20528366(35.26%)	20406630(35.06%)
CG3	59940624	47099978(78.58%)	42724925(71.28%)	4375053(7.30%)	22401622(37.37%)	22316850(37.23%)
CW1	89205520	60424721(67.74%)	54303433(60.87%)	6121288(6.86%)	28602353(32.06%)	28571807(32.03%)
CW2	73325390	54193072(73.91%)	47988486(65.45%)	6204586(8.46%)	25479683(34.75%)	25383354(34.62%)
CW3	82021138	59631700(72.70%)	53825598(65.62%)	5806102(7.08%)	28284423(34.48%)	28381940(34.60%)

注：Total Reads：Clean Reads 数目，按单端计；Mapped Reads：比对到参考基因组上的 Reads 数目及在 Clean Reads 中占的百分比；Uniq Mapped Reads：比对到参考基因组唯一位置的 Reads 数目及在 Clean Reads 中占的百分比；Multiple Mapped Reads：比对到参考基因组多个位置的数目及在 Clean Reads 中占的百分比；Reads Map to '+'：比对到参考基因组正链的 Reads 数目及在 Clean Reads 中占的百分比。Reads Map to '-'：比对到参考基因组负链的 Reads 数目及在 Clean Reads 中占的百分比。

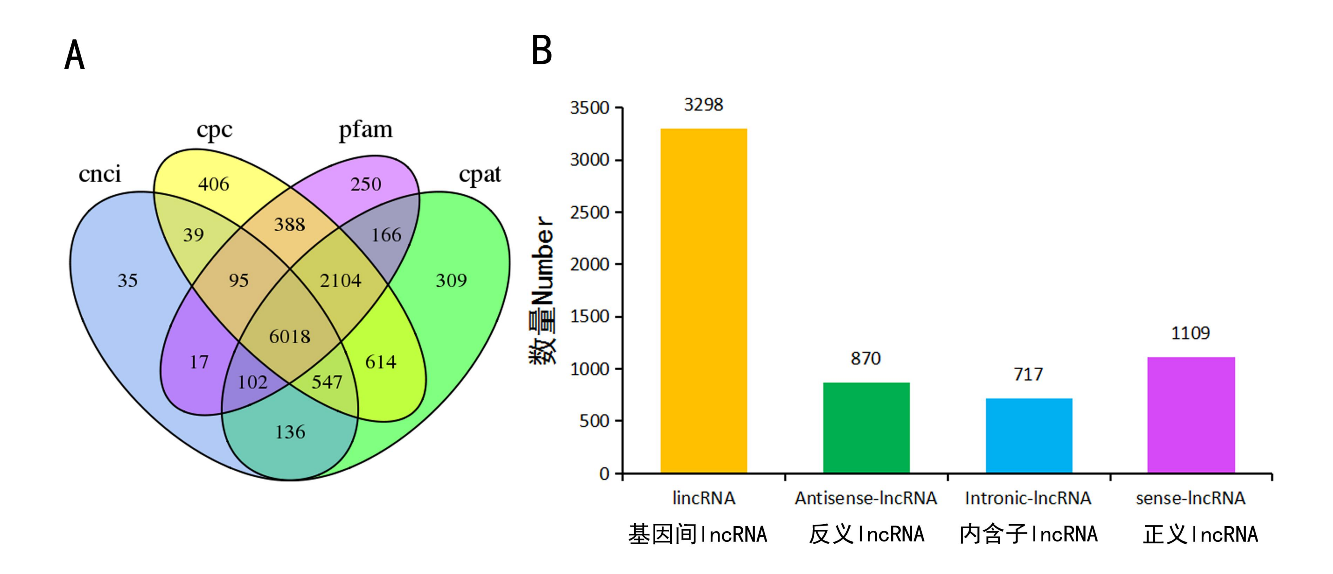
Note: Total Reads: Number of Clean Reads, calculated on a single end;Mapped Reads: the number of Reads compared to the reference genome and the percentage of Clean Reads; Uniq Mapped Reads: The number of Reads compared to the unique location in the reference genome and the percentage of Clean Reads; Multiple Mapped Reads: The number of Reads compared to multiple locations in the reference genome and the percentage of Clean Reads;Reads Map to '+': The number of Reads compared to the positive strand in the reference genome and the percentage of Clean Reads.Reads Map to '-': The number of Reads compared to the negative strand in the reference genome and the percentage of Clean Reads.

2.2 红苞凤梨 lncRNAs 的鉴定

对测序样品 cuffiinks 的拼接结果第一步经过 cuffcompare 软件分析合并，筛选出转录本长度>=200bp，外显子个数>=2，FPKM ≥ 0.1 的转录本，根据已知 mRNA 数据库，进行比较过滤 mRNA，然后通过 CNCI、CPC、CPAT 和 Pfam 四个软件进行蛋白编码潜能筛选，最终鉴定得到 6,018 条 lncRNAs，其中新 lncRNA5689 条（图 2）。比利用 Iso-seq 二代测序数据鉴定到的 lncRNAs 数量增加了约 70%（蔺珍，2019），极大地提高了红苞凤梨 lncRNA 的数据信息量，为进一步研究红苞凤梨的非编码 RNA 调控机理提供了数据基础。

鉴定得到的 6,018 个 lncRNAs 中，包括了 3,298 个基因间 lncRNAs（lincRNAs），717 个反义 lncRNAs（antisense lncRNAs），890 个内含子 lncRNAs（intronic lncRNAs）和 1,109 个正义 lncRNAs（sense lncRNAs）（图 3：B）。与二代测序分析鉴定结果相比，lincRNA

所在比例显著提高，由 17%增加到 55%。而 sense lncRNA 所占比例显著下降，由 76%下降到 18.5%（蔺珍，2019）。



注：**A**.四种鉴定方法得到的红苞凤梨 lncRNAs 数量 **B**. 红苞凤梨 lncRNAs 不同类型的分布

Note: **A**. Number of lncRNA of *Ananas comosus* var. *bracteatus* obtained by four identification methods

B. Distribution of different types of lncRNAs of *Ananas comosus* var. *bracteatus*

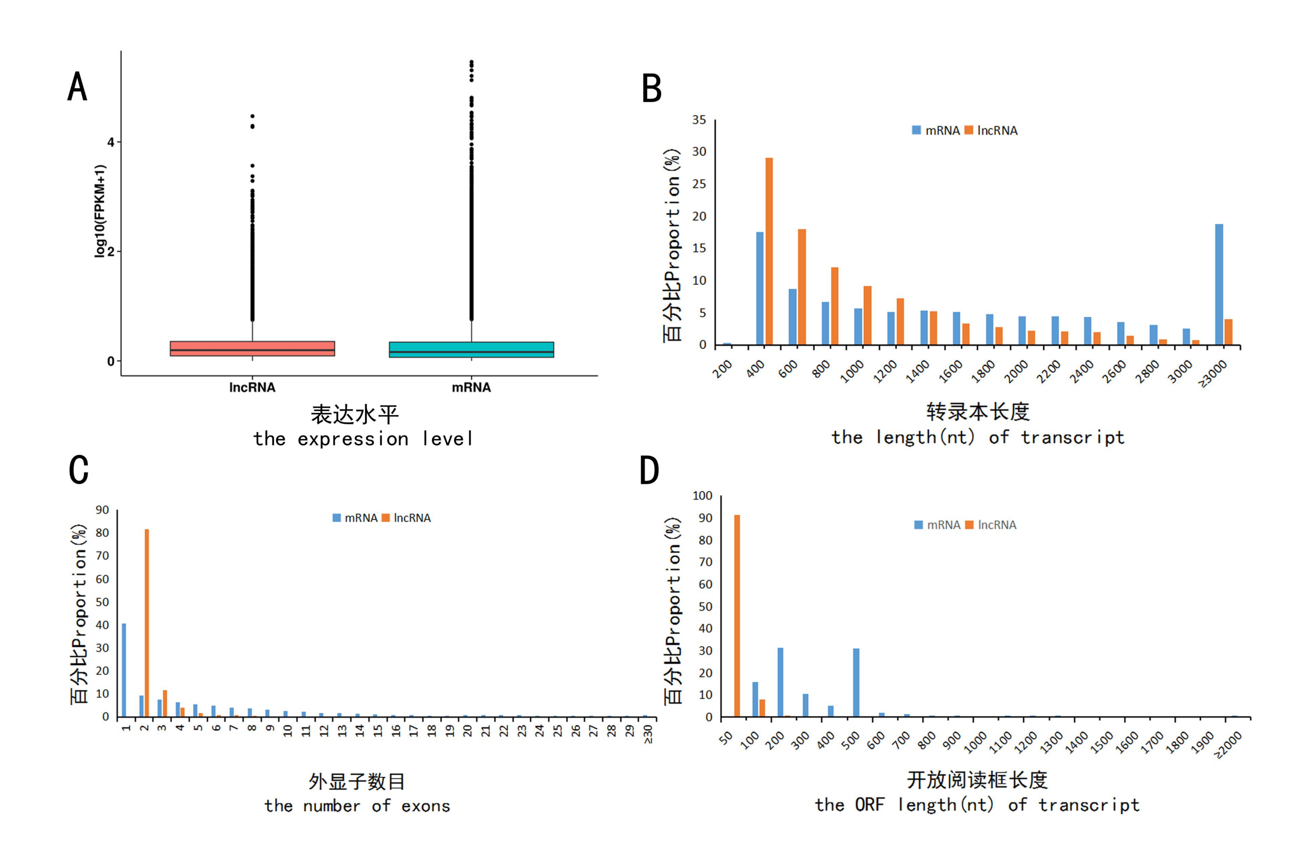
图 2 lncRNA 的鉴定及类型分布

Fig. 2 Identification and type distribution of lncRNA

2.3 红苞凤梨 lncRNA 结构分析

为了进一步红苞凤梨 lncRNA 的结构特点，将 lncRNA 与蛋白质编码 RNA 在整体表达水平、序列长度分布、外显子数目分布以及开放阅读框长度分布情况进行了比较分析（图 3）。分析结果表明，在总体表达水平上，mRNA 的表达丰度高于 lncRNA 的表达丰度（图 3：A）。在转录本的长度分布上，lncRNA 在 400~1200 nt 区间分布比例高于 mRNA（图 3：B），二代测序结果中则是表现在 400-600 nt 和 1400-1600 nt 区间（蔺珍，2019）；而在转录本长度>1600 nt 区间，lncRNA 分布的比例显著小于 mRNA，尤其在>=3 000 nt 区间（图 3：B）。lncRNA 中的外显子数量总体少于 mRNA，约 82%的 lncRNA 只含有 2 个外显子（图 3：C）。而二代测序分析中有 41.80%的 lncRNAs 只有 2 个外显子，而 31.62%的 mRNAs 的外显子数超过 5 个（蔺珍，2019）。lncRNA 的开放阅读框长度总体上也短于 mRNA。约 99%的 lncRNA 开放阅读框长度<=100 nt（图 3：D），而二代测序分析中 66%的 lncRNAs 的开放

阅读框在 0-100 nt 之间（蔺珍，2019）。



注：**A**. lncRNA 和 mRNA 的表达水平比较。**B**. lncRNA 和 mRNA 的转录本长度分布比较。**C**. lncRNA 和 mRNA 外显子数量分布比较。**D**.lncRNA 和 mRNA 的 ORF 长度分布比较。

Note: **A**. Comparison of the expression levels between lncRNAs and mRNAs. **B**. Comparison of transcript length distributions between lncRNA and mRNA. **C**. Comparison of the number of exons distributions between lncRNA and mRNA. **D**. Comparison of the ORF lengths distributions between lncRNA and mRNA.

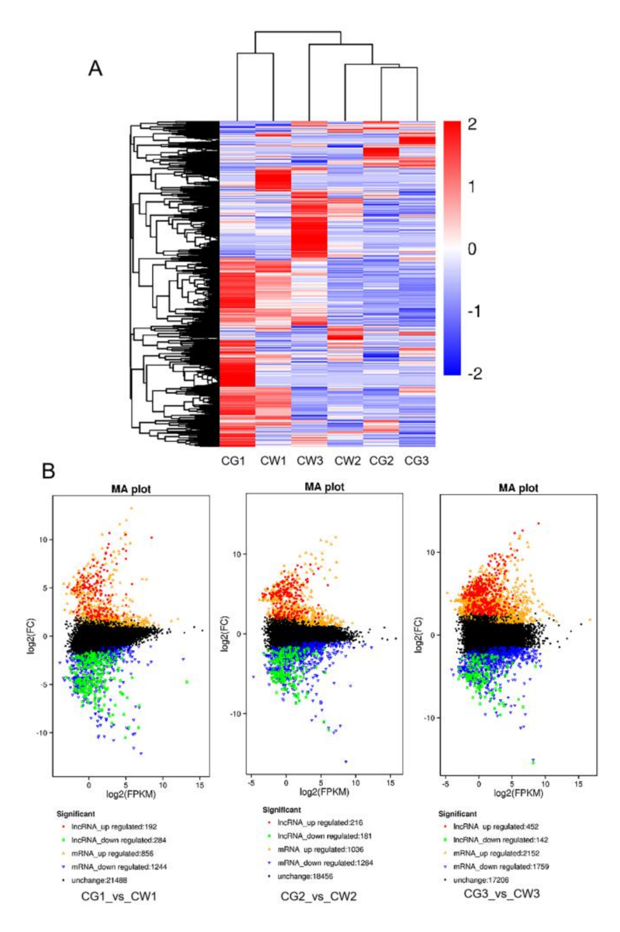
图 3 红苞凤梨 lncRNA 与 mRNA 的比较分析

Fig. 3 Comparative analysis of lncRNA and mRNA in *Ananas comosus var. bracteatus*

2.4 红苞凤梨 lncRNAs 的差异表达分析

以 Fold Change≥2.0 且 FDR<0.05 作为筛选标准，共鉴定得到了 1 710 个差异表达的 lncRNA。对筛选出的差异表达 lncRNA 进行层次聚类分析，聚类分析结果如图 4:A 所示。在未展叶阶段，全绿叶片和全白叶片中，大量差异表达的 lncRNA 表达丰度较高，在全白苗和全绿苗的发育到 4~5 叶期时，多数差异表达的 lncRNA 表达水平下降。而全白苗在第三发育阶段有部分差异表达 lncRNA 表达水平显著上调。在三个发育阶段全绿全白叶片间显著差

异表达的 lncRNA 可能是红苞凤梨嵌合性状形成的关键调控因子。



注：**A**.差异表达 lncRNA 聚类图。 图中颜色代表了 lncRNA 在样品中的表达量水平 log2(FPKM+1)。

B. 差异表达 lncRNA 和 mRNA 表达量的 MA 交互式图。差异表达基因 MA 图中每一个点代表一个基因。横坐标为两样品中表达量均值的对数值；纵坐标为两样品间基因表达量差异倍数的对数值。红色代表上调 lncRNA，绿色代表下调 lncRNA，橘色代表上调基因，蓝色代表下调基因，黑色的点代表表达差异不显著的基因。

Note： **A**. Differentially expressed lncRNA clustering maps.The color in the figure represents the level of expression of lncRNA in the sample log2 (FPKM+1). **B**. MA interactive map of differentially expressed lncRNA and mRNA expression Each point in the MA map of the differentially expressed gene represents a gene. The abscissa is the logarithm of the mean value of the expression in the two samples; the ordinate is the logarithm of the difference in the gene expression between the two samples. Red represents up-regulation of lncRNA, green represents down-regulation of lncRNA, orange represents up-regulated genes, blue represents down-regulated genes, and black dots represent genes with insignificant differences in expression.

图 4 差异表达 lncRNA 分析。

Fig. 4 Differentially expressed lncRNA analysis

同一发育阶段的全绿叶片与全白叶片差异表达的 lncRNA 数量和 mRNA 数量如图 4B 所示。在未展叶阶段，差异表达 lncRNA 共 476 个，其中在全白叶片中上调表达的 lncRNA192 个，占比约 40%；差异表达的 mRNA 共 3911 个，其中上调表达的 mRNA2152 个，占比 55%。在 4~5 叶阶段，差异表达 lncRNA 共 397 个，其中在全白叶片中上调表达的 lncRNA216 个，占比约 54%；差异表达的 mRNA 共 2300 个，其中上调表达的 1036 个，占比 45%。在 10~12 片叶阶段，差异表达 lncRNA 共 594 个，其中在全白叶片中上调表达的 lncRNA452 个，占比约 76%；差异表达的 mRNA 共 2100 个，其中上调表达的 856 个，占比约 41%。可以看出，随着植株的生长，lncRNA 在全绿植株与全白植株间的差异表达越发显著，在全白植株中上调表达的 lncRNA 显著增加，lncRNA 的差异表达可能在绿、白叶色的差异形成过程中起着重要的调控作用。而这个过程中，差异表达的 mRNA 数量在减少，上调表达的基因占比也在减少。

2.5 红苞凤梨 LncRNA 靶基因预测

LncRNA 调控其靶基因的方式有两种，分为 Cis 作用和 trans 作用。根据 Cis 作用，我们将 lncRNA 100kb 范围内的邻近蛋白质编码基因为其靶基因，经过分析预测，5441 个 lncRNAs 通过 Cis 作用方式预测到靶基因；trans 作用是 lncRNA 与 mRNA 由于碱基互补配对而产生作用，LncTar（LI et al.，2015）正是利用 LncRNA 和 mRNA 之间存在的互补配对关系进行预测，通过计算配对位点自由能和标准化自由能，标准化自由能阈值以下的则认为是 LncRNA 的靶基因。1544 个 lncRNAs 通过 Trans 方式预测到靶基因。靶基因的预测可以帮助理解 lncRNA 的功能，揭示 lncRNA 在红苞凤梨生长发育过程中的调控作用。

2.6 红苞凤梨差异表达 lncRNA 靶基因功能注释和富集分析

2.6.1 差异表达 lncRNA 顺式靶基因功能注释和富集分析

对差异表达 lncRNA 的顺式靶基因进行 COG、GO 、KEGG、KOG、NR 和 Swiss-Prot 功能富集分析。富集分析结果如表 2 所示。

表 2 注释的差异表达 lncRNA 顺式靶基因数量统计表

Table 2 Statstical of the number of cis-target genes of differentially expressed lncRNA

组别	注释个数	COG	GO	KEGG	KOG	NR	Swiss-Prot

DEG Set	Annotated	数据库	数据库	数据库	数据库	数据库	数据库
CG1_vs_CW1	1878	721	1219	719	1131	1875	1298
CG2_vs_CW2	1667	632	1093	657	1013	1666	1179
CG3_vs_CW3	2263	885	1525	925	1394	2260	1619
CG1_vs_CG2	1987	761	1342	779	1196	1985	1393
CG1_vs_CG3	2283	861	1542	899	1350	2280	1596
CW1_vs_CW2	2124	817	1421	845	1287	2123	1515
CW1_vs_CW3	2284	854	1558	908	1376	2281	1635

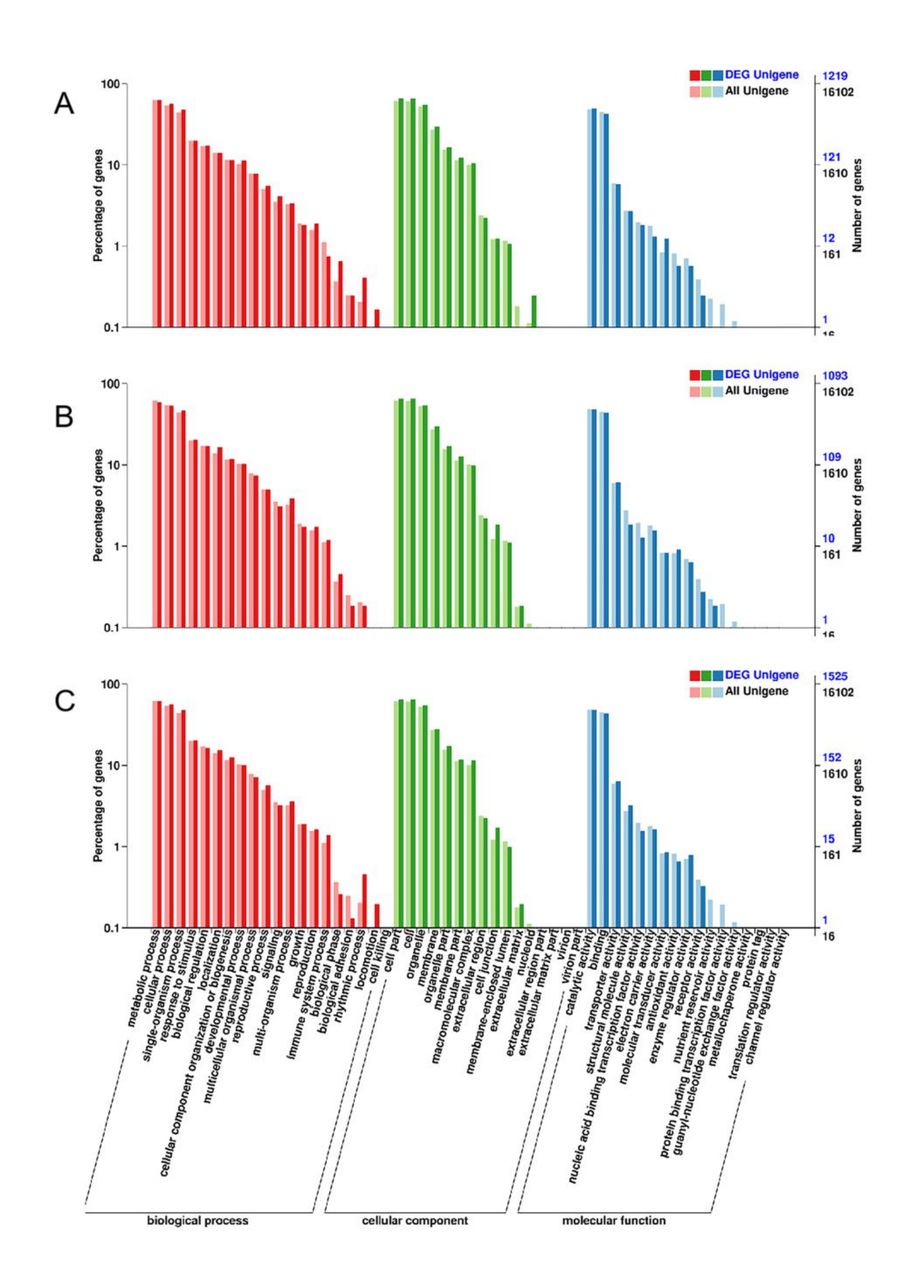
DEG Set	Annotated	COG	GO	KEGG	KOG	NR	Swiss-Prot
CG1_vs_CW1	1878	721	1219	719	1131	1875	1298
CG2_vs_CW2	1667	632	1093	657	1013	1666	1179
CG3_vs_CW3	2263	885	1525	925	1394	2260	1619
CG1_vs_CG2	1987	761	1342	779	1196	1985	1393
CG1_vs_CG3	2283	861	1542	899	1350	2280	1596
CW1_vs_CW2	2124	817	1421	845	1287	2123	1515
CW1_vs_CW3	2284	854	1558	908	1376	2281	1635

注：COG 数据库：蛋白质直系同源基因簇（Clusters of Orthologous Groups）；GO 数据库：基因本体（gene ontology）；KEGG 数据库：京都基因与基因组百科全书（Kyoto Encyclopedia of Genes and Genomes）；KOG 数据库：真核生物直系同源基因簇（Clusters of orthologous groups for eukaryotic complete genomes）；NR 数据库：非冗余蛋白库（Non-Redundant Protein Sequence Database）；Swiss-Prot 数据库：含有详细注释内容的的蛋白序列库。

Note: COG: Clusters of Orthologous Groups; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; KOG: Clusters of orthologous groups for eukaryotic complete genomes; NR: Non-Redundant Protein Sequence Database; Swiss-Prot is a protein sequence database with detailed annotation content.

GO(gene ontology)数据库是一个结构化的标准生物学注释系统，在 GO 分析中，基因注释在三个层次上，即：生物学途径（biological process），分子功能（molecular function）和细胞组件（cellular component）。从图 5 可以看出，在未展叶阶段，差异表达 lncRNA 的靶基因主要富集在生物学途径中的 biological phases，rhythmic process 和 locomotion，细胞组

件方面的 extracellular matrix 和 nucleoid，分子功能方面的 nutrient reservoir activity、protein binding transcription factor activity、guanyl-nucleotide exchange factor activity。在 4~5 片叶阶段，差异表达 lncRNA 的靶基因主要富集在细胞组件的 nucleotide 和分子功能的 protein binding transcription factor activity、guanyl-nucleotide exchange factor activity。在 10~12 片叶阶段，差异表达 lncRNA 的靶基因主要富集在生物学途径中的 biological adhesion、rhythmic process 和 locomotion，分子功能的 nutrient reservoir activity、protein binding transcription factor activity、guanyl-nucleotide exchange factor activity。



注：横坐标为 GO 分类，纵坐标左边为基因数目所占百分比，右边为基因数目。此图展示的是在差异表达基因背景和全部基因背景下 GO 各二级功能的基因富集情况，体现两个背景下各二级功能的地位，具有明

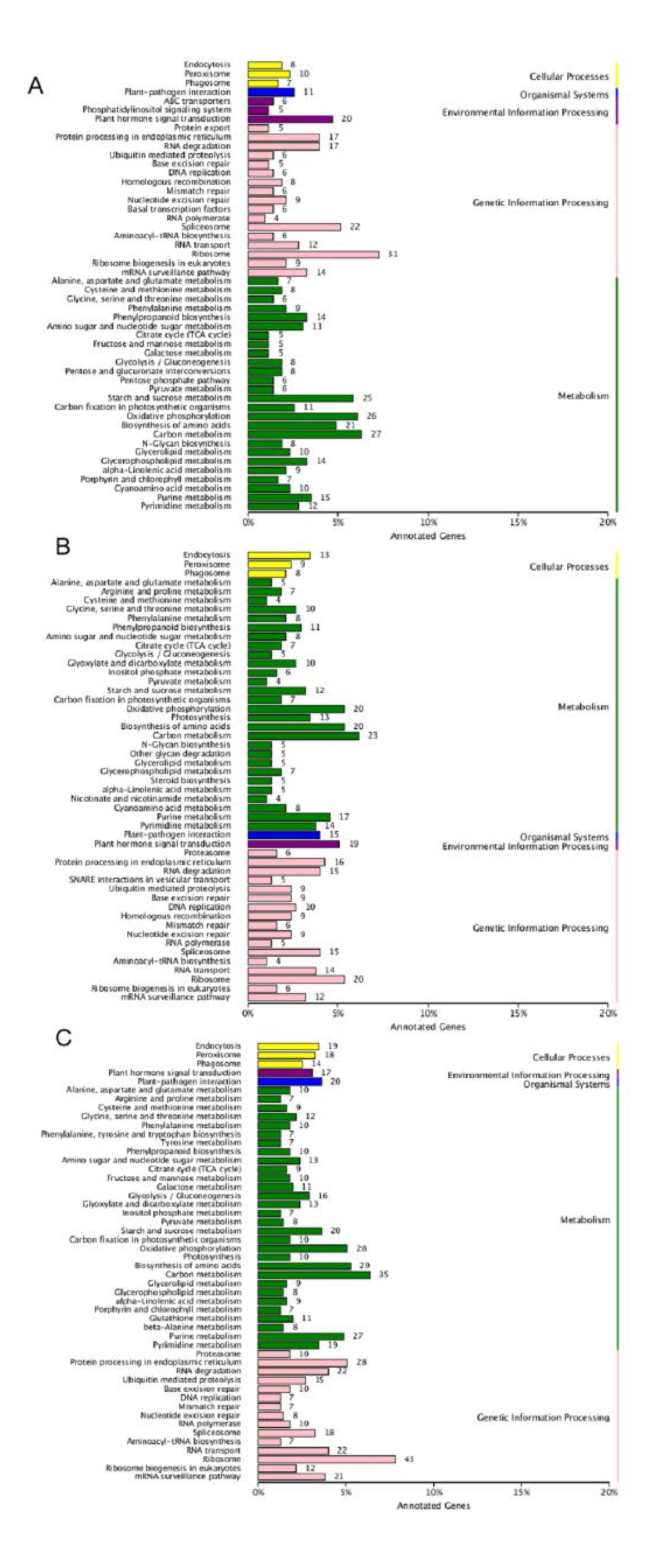
显比例差异的二级功能说明差异表达基因与全部基因的富集趋势不同，可以重点分析此功能是否与差异相关。

Note: The abscissa is the GO classification, the left side of the ordinate is the percentage of the number of genes, and the right side is the number of genes. This figure shows the enrichment of the secondary functions of GO in the background of differentially expressed genes and the background of all genes, reflecting the status of each secondary function in two contexts, and the secondary function with significant proportional differences indicates that the differentially expressed genes are different from the enrichment trend of all genes.it is important to analyze whether this function is related to differences.

图 5 差异表达基因 GO 注释分类统计图

Fig. 5 Differential expression gene GO annotation classification chart

在生物体内，不同的基因产物相互协调以执行生物学功能，在 GO 分析基础上，对差异表达 lncRNA 反式靶基因的信号通路注释分析能够进一步解读基因的功能。由图 7 可看出，在未展叶期，全绿叶片与全白叶片间的差异表达 lncRNA 的靶基因主要富集在核糖体、碳代谢、氧化磷酸化、淀粉和糖代谢、氨基酸代谢、脂类代谢等这些基础代谢和植物激素信号转录通路上。同时，在卟啉和叶绿素代谢途径中也富集了 7 个差异表达基因。在 4~5 片叶阶段，差异表达 lncRNA 的靶基因主要富集在碳代谢、氧化磷酸化、氨基酸代谢、核糖体这些基础代谢上。同时，也富集在植物激素信号转导、嘌呤代谢、内质网蛋白过程、RNA 降解、光合作用、淀粉和糖代谢方面。在 10~12 叶阶段，差异表达 lncRNA 的靶基因主要富集在核糖体、碳代谢、氨基酸代谢、氧化磷酸化、内质网蛋白过程、嘌呤代谢、RNA 降解、RNA 转运等方面，在卟啉和叶绿素合成代谢途径、光合作用途径中也富集了差异表达基因。与二代测序结果（蔺珍，2019）对比发现，联合测序分析增加的靶基因主要富集在碳代谢、氨基酸代谢、氧化磷酸化、RNA 降解等通路上。差异表达基因的富集结果充分说明 lncRNA 参与了红苞凤梨叶片色素的合成、光合作用、物质代谢、生长发育调控等生理过程的调控。


A
Endocytosis 8
Peroxisome 10
Phagosome 7
Plant-pathogen interaction 11
ABC transporters 6
Phosphatidylinositol signaling system 5
Plant hormone signal transduction 20
Protein export 5
Protein processing in endoplasmic reticulum 17
RNA degradation 17
Ubiquitin mediated proteolysis 6
Base excision repair 5
DNA replication 6
Homologous recombination 8
Mismatch repair 6
Nucleotide excision repair 9
Basal transcription factors 6
RNA polymerase 4
Spliceosome 22
Aminoacyl-tRNA biosynthesis 6
RNA transport 12
Ribosome 31
Ribosome biogenesis in eukaryotes 9
mRNA surveillance pathway 14
Alanine, aspartate and glutamate metabolism 7
Cysteine and methionine metabolism 8
Glycine, serine and threonine metabolism 6
Phenylalanine metabolism 9
Phenylpropanoid biosynthesis 14
Amino sugar and nucleotide sugar metabolism 13
Citrate cycle (TCA cycle) 5
Fructose and mannose metabolism 5
Galactose metabolism 5
Glycolysis / Gluconeogenesis 8
Pentose and glucuronate interconversions 8
Pentose phosphate pathway 6
Pyruvate metabolism 6
Starch and sucrose metabolism 25
Carbon fixation in photosynthetic organisms 11
Oxidative phosphorylation 26
Biosynthesis of amino acids 21
Carbon metabolism 27
N-Glycan biosynthesis 8
Glycerolipid metabolism 10
Glycerophospholipid metabolism 14
alpha-Linolenic acid metabolism 9
Porphyrin and chlorophyll metabolism 7
Cyanoamino acid metabolism 10
Purine metabolism 15
Pyrimidine metabolism 12
Cellular Processes
Organismal Systems
Environmental Information Processing
Genetic Information Processing
Metabolism
0% 5% 10% 15% 20%
Annotated Genes
B
Endocytosis 13
Peroxisome 9
Phagosome 8
Alanine, aspartate and glutamate metabolism 5
Arginine and proline metabolism 7
Cysteine and methionine metabolism 4
Glycine, serine and threonine metabolism 10
Phenylalanine metabolism 8
Phenylpropanoid biosynthesis 11
Amino sugar and nucleotide sugar metabolism 8
Citrate cycle (TCA cycle) 7
Glycolysis / Gluconeogenesis 5
Glyoxylate and dicarboxylate metabolism 10
Inositol phosphate metabolism 6
Pyruvate metabolism 4
Starch and sucrose metabolism 12
Carbon fixation in photosynthetic organisms 7
Oxidative phosphorylation 20
Photosynthesis 13
Biosynthesis of amino acids 20
Carbon metabolism 23
N-Glycan biosynthesis 5
Other glycan degradation 5
Glycerolipid metabolism 5
Glycerophospholipid metabolism 7
Steroid biosynthesis 5
alpha-Linolenic acid metabolism 5
Nicotinate and nicotinamide metabolism 4
Cyanoamino acid metabolism 8
Purine metabolism 17
Pyrimidine metabolism 14
Plant-pathogen interaction 15
Plant hormone signal transduction 19
Proteasome 6
Protein processing in endoplasmic reticulum 16
RNA degradation 15
SNARE interactions in vesicular transport 5
Ubiquitin mediated proteolysis 9
Base excision repair 9
DNA replication 10
Homologous recombination 9
Mismatch repair 6
Nucleotide excision repair 9
RNA polymerase 5
Spliceosome 15
Aminoacyl-tRNA biosynthesis 4
RNA transport 14
Ribosome 20
Ribosome biogenesis in eukaryotes 6
mRNA surveillance pathway 12
Cellular Processes
Metabolism
Organismal Systems
Environmental Information Processing
Genetic Information Processing
0% 5% 10% 15% 20%
Annotated Genes
C
Endocytosis 19
Peroxisome 18
Phagosome 14
Plant hormone signal transduction 17
Plant-pathogen interaction 20
Alanine, aspartate and glutamate metabolism 10
Arginine and proline metabolism 7
Cysteine and methionine metabolism 9
Glycine, serine and threonine metabolism 12
Phenylalanine metabolism 10
Phenylalanine, tyrosine and tryptophan biosynthesis 7
Tyrosine metabolism 7
Phenylpropanoid biosynthesis 10
Amino sugar and nucleotide sugar metabolism 13
Citrate cycle (TCA cycle) 9
Fructose and mannose metabolism 10
Galactose metabolism 11
Glycolysis / Gluconeogenesis 16
Glyoxylate and dicarboxylate metabolism 13
Inositol phosphate metabolism 7
Pyruvate metabolism 8
Starch and sucrose metabolism 20
Carbon fixation in photosynthetic organisms 10
Oxidative phosphorylation 28
Photosynthesis 10
Biosynthesis of amino acids 29
Carbon metabolism 35
Glycerolipid metabolism 9
Glycerophospholipid metabolism 8
alpha-Linolenic acid metabolism 9
Porphyrin and chlorophyll metabolism 7
Glutathione metabolism 11
beta-Alanine metabolism 8
Purine metabolism 27
Pyrimidine metabolism 19
Proteasome 10
Protein processing in endoplasmic reticulum 28
RNA degradation 22
Ubiquitin mediated proteolysis 15
Base excision repair 10
DNA replication 7
Mismatch repair 7
Nucleotide excision repair 8
RNA polymerase 10
Spliceosome 18
Aminoacyl-tRNA biosynthesis 7
RNA transport 22
Ribosome 43
Ribosome biogenesis in eukaryotes 12
mRNA surveillance pathway 21
Cellular Processes
Environmental Information Processing
Organismal Systems
Metabolism
Genetic Information Processing
0% 5% 10% 15% 20%
Annotated Genes

注：**A**.未展叶阶段全绿全白叶片间（CG1_vs_CW1）差异表达 lncRNA 靶基因 KEGG 富集分析。

B.4~5 片叶阶段全绿全白叶片间（CG2_vs_CW2）差异表达 lncRNA 靶基因 KEGG 富集分析。

C.10~12 片叶阶段全绿全白叶片间（CG3_vs_CW3）差异表达 lncRNA 靶基因 KEGG 富集分析。

Note: **A**. KEGG enrichment analysis of target genes of differentially expressed lncRNAs in the unexpanded leaf stage between complete green and complete white leaves (CG1_vs_CW1). **B.** KEGG enrichment analysis of target genes of differentially expressed lncRNAs in 4-5 leaf stage between complete green and complete white leaves (CG2_vs_CW2). **C**. KEGG enrichment analysis of target genes of differentially expressed lncRNAs in 10-12 leaf stage between complete green and complete white leaves (CG3_vs_CW3).

图 6 差异表达 lncRNA 顺式靶基因 KEGG 分类图

Fig. 6 Differentially expressed lncRNA cis-target gene KEGG Classification map

2.6.2 差异表达 lncRNA 反式靶基因功能注释和富集分析

对差异表达 lncRNA 反式靶基因进行 COG、GO 、KEGG、KOG、NR 和 Swiss-Prot 功能富集分析。富集分析结果如表 3 所示。

表 3 注释的差异表达 lncRNA 反式靶基因数量统计表

Tab3 Statistical table of the number of trans-target genes of Annotated differentially expressed lncRNA

组别 DEG Set	注释个数 Annotated	COG 数据库	GO 数据库	KEGG 数据库	KOG 数据库	NR 数据库	Swiss-Prot 数据库
CG1_vs_CW1	14	7	9	4	6	14	5
CG2_vs_CW2	29	7	20	8	14	29	14
CG3_vs_CW3	29	7	19	2	16	29	10
CG1_vs_CG2	16	7	11	3	6	16	7
CG1_vs_CG3	24	8	15	5	8	24	8
CW1_vs_CW2	13	2	8	4	5	13	8
CW1_vs_CW3	22	5	14	1	11	22	12

DEG Set	Annotated	COG	GO	KEGG	KOG	NR	Swiss-Prot
CG1_vs_CW1	14	7	9	4	6	14	5
CG2_vs_CW2	29	7	20	8	14	29	14
CG3_vs_CW3	29	7	19	2	16	29	10
CG1_vs_CG2	16	7	11	3	6	16	7
CG1_vs_CG3	24	8	15	5	8	24	8
CW1_vs_CW2	13	2	8	4	5	13	8
CW1_vs_CW3	22	5	14	1	11	22	12

注：COG 数据库：蛋白质直系同源基因簇（Clusters of Orthologous Groups）；GO 数据库：基因本体（gene ontology）；KEGG 数据库：京都基因与基因组百科全书（Kyoto Encyclopedia of Genes and Genomes）；KOG 数据库：真核生物直系同源完整基因簇（Clusters of orthologous groups for eukaryotic complete genomes）；NR 数据库：非冗余蛋白库（Non-Redundant Protein Sequence Database）；Swiss-Prot 数据库：含有详细注释内容的的蛋白序列库。

Note: COG: Clusters of Orthologous Groups; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; KOG: Clusters of orthologous groups for eukaryotic complete genomes; NR: Non-Redundant Protein Sequence Database; Swiss-Prot is a protein sequence database with detailed annotation content.

功能注释和富集分析结果表明，差异表达 lncRNA 反式靶基因功能注释到的基因数量不多,在叶片发育过程中，全绿、全白叶片间差异表达 lncRNA 顺式靶基因主要富集在核糖体、碳代谢、氨基酸的生物合成、氧化磷酸化、内质网中的蛋白质加工以及淀粉与蔗糖代谢等方面；而差异表达 lncRNA 反式靶基因主要富集在 TCA 循环、淀粉和糖代谢、氨基糖和核糖代谢、RNA 降解、氨基酸代谢等方面。两类 lncRNA 在不同的生理代谢过程中发挥着调控作用。

3. 讨论

植物叶色嵌合体嵌合性状明显，嵌合方式多样且易于观察，是研究植物生长发育以及遗传育种的优良材料。对植物叶色嵌合体的研究，已经成为嵌合体研究的重要方向，深入研究

植物嵌合体形成机理，对于植物生长发育过程中细胞间的相互作用、植物嵌合性状稳定繁殖以及植物遗传育种具有十分重要的意义。红苞凤梨叶花果具有艳丽的颜色，是重要的新型观赏植物，是研究叶色镶嵌形成机理的理想材料。叶肉细胞的白化突变是金边嵌合叶色形成的基础和关键环节，是多基因协同作用的结果。嵌合性状的形成与稳定，与基因表达的有序调控是密切相关的。前期研究表明，转录后调控在红苞凤梨金边嵌合性状形成过程中具有重要作用，而 lncRNA 可在表观遗传调控、转录调控、转录后调控等水平发挥功能，构建红苞凤梨 lncRNA 表达谱，揭示 lncRNA 的调控机制对于弄清红苞凤梨金边嵌合性状形成机理具有重要意义。由于红苞凤梨基因组信息的缺乏，lncRNA 的鉴定分析以二代转录组测序数据为基础，以菠萝基因组作为参考基因组进行分析。红苞凤梨与菠萝为同属同种不同变种植物，亲缘关系很近（LEAL&COPPENS，2003），采用菠萝基因组数据作为参考基因组能有效鉴定红苞凤梨的 lncRNA，为红苞凤梨的表观遗传调控研究提供基础数据。而二代测序技术由于其读长短，不能提供完整的转录本（KOREN et al.，2012），很难正确预测基因的结构（COGHLAN et al.，2008）。SMRT 全长转录组测序克服了二代测序读长短的缺点，是研究基因结构、基因功能和比较基因组学的基本方法（LUO et al.，2017；SHARON et al.，2013）。本研究采用 SMRT 全长转录组测序数据和二代测序数据混合组装修正后进行 lncRNA 分析，以提高 lncRNA 分析的准确性。

经混合组装校正后的 clean reads 比对到菠萝参考基因组，比对效率达到 67.74%~78.58%，比二代数据分析提高了约 5%，有效地提高了长链非编码数据的利用率（蔺珍，2019）。经 CPC 分析、CNCI 分析、CPAT 分析、pfam 蛋白结构域分析，共鉴定得到 6，018 个 lncRNA，比之前鉴定得到的 lncRNA 数量提高了约 70%，极大地丰富了红苞凤梨 lncRNA 数据库，为红苞凤梨表观遗传调控的研究提供了基础数据。在鉴定得到的 6018 个 lncRNA 中，最多的是 lincRNA，占比约 55%，比之前结果提高了约 2 倍。而 sense lncRNA 的数量极显著下降，由之前的 75%下降到 18.5%（蔺珍，2019）。本次分析结果中 4 种 lncRNA 的数量分布与玉米（*Zea mays*）（WANG et al.，2016）基本相似，修正了之前分析结果中，sense lncRNA 比例太高的情况。SMRT 全长转录组测序数据与 lncRNA 二代转录组测序数据的联合分析，有效地提高了 lncRNA 测序数据的利用效率，很大程度上弥补了以近缘种菠萝基因组作为参考基因组鉴定 lncRNA 所带来的缺陷，鉴定得到的 lncRNA 数量极显著提高，丰富了红苞凤梨 lncRNA 数据库信息。联合分析结果中，修正了 4 种 lncRNA 数量分布的异常，提高了 lncRNA 鉴定结果的准确性，有效地提高了后期相关研究的可行性和可靠性。lncRNA 和 mRNA 序列结构对比分析表明，与编码基因相比，lncRNA 具有表达丰度较低、序列长度较短，外显子

数目较少，ORF 长度较短等结构特点。这与斑马鱼（*Brachydanio rerio*）（高霄霄，2017）、臭橘（*Poncirus trifoliata*）（WANG et al.，2017）、杨树（*Populus*）（田净净，2016）的分析结果一致，说明 lncRNA 的这些结构特征在生物中具有普遍性。

lncRNA 靶基因的预测以及靶基因的功能注释和富集分析，是研究 lncRNA 功能的重要途径（蔺珍，2019）。本研究中，GO 和 KEGG 的联合分析表明，在所研究的叶片发育的三个阶段中，差异表达 lncRNA 的靶基因注释到了叶片生长发育的多个方面，包括了碳代谢、氨基酸代谢、脂代谢、核糖体、淀粉和糖代谢等这些基础代谢方面，也包括了植物激素信号转导等调控机制方面。在未展叶期，差异表达 lncRNA 的靶基因注释到了卟啉和叶绿素代谢途径，说明在叶片发育的早期，全绿全白叶片的叶绿素合成代谢就出现了差异，导致了叶片颜色的差异。在 4~5 叶期和 10~12 叶期，光合作用途径富集了差异表达基因，说明由于叶片的失绿白化，光合作用受到抑制，相关基因差异表达。同时，由于光合作用能力的差异，全绿全白叶片在各个生理代谢方面都表现出差异，在很多基础代谢、调控途径、核酸代谢等方面都富集了差异表达基因。这些基因表达模式说明，lncRNA 对靶基因表达的调控作用可能在红苞凤梨叶片发育早期叶绿素合成代谢差异以及后期光合作用差异和其他生理代谢差异方面都起着重要的调控作用。本研究中鉴定出的相关差异表达 lncRNA 信息，为进一步研究 lncRNA 对靶基因的调控作用及靶基因在红苞凤梨叶片失绿白化突变机制，全面揭示红苞凤梨金边嵌合体形成的机理提供了重要的数据基础。

参考文献

CAO L, 2011. A Study on In Vitro Culture of Chimera Cultivars of Ananas Bracteatus Schultes and Their Stability of Chimeric Traits[D]. South China Agricultural University：1-57. [曹莉，2011.红苞凤梨嵌合体品种离体培养及其稳定性的研究[D].华南农业大学：1-57.]

CSORBA T, QUESTA JI, SUN Q, et al.， 2014. Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization[J]. Proc Nat Acad Sci USA, 111(45):16160-16165.

COGHLAN A, FIEDLER TJ, MCKAY SJ, et al.， 2008. nGASP—the nematode genome annotation assessment project[J]. BMC Bioinformatics, 9:549DOI 10.1186/1471-2105-9-549.

DING J, LU Q, OUYANG Y, et al.，2012. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice[J]. Proc Nat Acad Sci USA, 109(7):2654-2659.

ENGLISH A C, RICHARDS S, HAN Y, et al. ，2012. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology[J]. Plos One, 7(11):e47768.

FLUSBERG B A, WEBSTER D R, LEE J H, et al. ，2010. Direct detection of DNA methylation during single-molecule, real-time sequencing[J]. Nat Methods, 7(6):461-465.

FINN R D, BATEMAN A, CLEMENTS J, et al. ，2014. Pfam: the protein families database[J]. Nucl Acid Res, 42(Database issue): D222-D230.

GUO F, WANG D, WANG L, 2018. Progressive approach for SNP calling and haplotype assembly using single molecular sequencing data[J]. Bioinformatics, 34(12):2012-2018.

GAO XX, 2017. Screening and identification of long noncoding RNAs in thepubertal female goats[D]. Anhui Agricultural University: 1-39. [高霄霄，2017. 初情期雌性山羊 lncRNA 的筛选与鉴定[D].安徽农业大学: 1-39.]

HACKL T, HEDRICH R, SCHULTZ J, et al. ，2014. proovread: large-scale high-accuracy PacBio correction through iterative short read consensus[J]. Bioinformatics, 30(21):3004-3011.

JOHNSON R, 2012. Long non-coding RNAs in Huntington's disease neurodegeneration[J]. Neurobiol Disease, 46(2):245-254.

KOREN S, WALENZ B P, BERLIN K, et al. ，2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation[J]. Genome Res, 27(5):722-736.

KOREN S, SCHATZ M C, WALENZ B P, et al. ，2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads[J]. Nat Biotechnol, 30(7):693-700.

KIM D, PERTEA G, TRAPNELL C, et al. ，2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions[J]. Genome Biol, 14(4):R36.

KONG L, ZHANG Y, YE Z Q, et al.，2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine[J]. Nucl Acid Res, 35(Web Server issue):W345-W349.

KOREN S, SCHATZ MC, WALENZ BP, MARTIN J, HOWARD JT, GANAPATHY G, 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads[J]. Nat Biotechnol, 30(7):693–700 DOI 10.1038/nbt.2280.

LEAL F, COPPENS DG, 2003. Morphology, anatomy and taxonomy[M]. Morphology, anatomy and taxonomy.

LIN Z, 2019. Identification of A. *comosus* var. *bracteatus* lncRNAs and functional verification of lncABCG11[D]. Ya'an: Sichuan Agricultural University： 1-113 [蔺珍，2019. 红苞凤梨 lncRNAs 的鉴定及 lncABCG11 的功能验证[D]. 雅安：四川农业大学：1-113.]

LI X , KANAKALA S , HE Y , et al. ， 2017. Physiological Characterization and Comparative Transcriptome Analysis of White and Green Leaves of *Ananas comosus* var. *bracteatus*[J]. Plos One, 12(1): e0169838.

LANGMEAD B, TRAPNELL C, POP M, et al. ，2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome[J]. Genome Biol, 10(3): R25.

LI J, MA W, ZENG P, et al. ， 2015. LncTar: a tool for predicting the RNA targets of long noncoding RNAs[J]. Brief Bioinform, 16(5): 806-812.

LUO YH, DING N, SHI X, WU YX, WANG RY, PEI LQ, XU RY, CHENG S, LIAN YY, GAO JY, WANG AM, TANG J, CAO QH, 2017. Generation and comparative analysis of full-length transcriptomes in sweetpotato and its putative wild ancestor *I. trifida*[J]. BioRxiv, https://doi.org/10.1101/112425

MA JC, 2018. Genome sequence of a widely cultivated poplar and its lnc RNAs response to salt stress[D]. Lanzhou： Lanzhou University：1-82. [马建超，2018. 新疆杨基因组及其 lncRNA 响应盐胁迫的研究[D]. 兰州：兰州大学：1-82.]

MA DN, ZHANG XT, WEI LF, et al. ， 2018. Benchmarking Hybrid Correction and Assembly Using Short Illumina Reads and Long Pac Bio Reads[J]. Genom Appl Biol, 37(04):1547-1555. [马东娜，张兴坦，魏柳锋，等，2018. 基因组二代测序数据与三代测序数据的混合校正和组装[J]. 基因组学与应用生物学, 37(04):1547-1555.]

MA J, XIANG YX, XIONG YY, et al. ， 2018. SMRT sequencing analysis reveals the full-length transcripts and alternativesplicing patterns in *Ananas comosus* var.*bracteatus*[J]. Peer J, 7：e7062

QIN T, ZHAO H, CUI P, et al. ， 2017. A Nucleus-Localized Long Non-Coding RNA Enhances Drought and Salt Stress Tolerance[J]. Plant Physiol, 175(3):1321-1336.

SMITH C C, WANG Q, CHIN C S, et al. ， 2012. Validation of ITD mutations in FLT3 as a therapeutic target in human acute myeloid leukaemia[J]. Nature, 485(7397):260-263.

SUN L, LUO H, BU D, et al. ， 2013. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts[J]. Nucl Acid Res, 41(17):e166.

SHARON D, TILGNER H, GRUBERT F, SNYDER M, 2013. A single-molecule long-readsurvey of the human transcriptome[J]. Nat Biotechnol, 31:1009–1014DOI 10.1038/nbt.2705.

TRAPNELL C, WILLIAMS B A, PERTEA G, et al. ， 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation[J]. Nat Biotechnol, 28(5):511-515.

TIAN JJ, 2016. The application of CRISPR/Cas9 system in the zebrafish gene editing[D]. Yangzhou University: 1-68. [田净净，2016CRISPR/Cas9 系统在斑马鱼基因编辑研究中的应用[D]. 扬州大学:1-68.]

WANG YF, SU WY, ZHANG L, et al. ，2018. Advances of Long Non-coding RNA in Plants[J]. Acta Botanica Boreali-Occidentalia Sinica, (3):582-588. [王艳芳，苏婉玉，张琳，等.，2018. 植物中 lncRNAs 的研究进展[J]. 西北植物学报, (3):582-588.]

WANG L, PARK H J, DASARI S, et al. ，2013. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model[J]. Nucl Acid Res, 41(6):e74.

WANG B , TSENG E , REGULSKI M , et al. ， 2016. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing[J]. Nat Comm, 7:11708.

WANG C, LIU S, ZHANG X, et al. ，2017. Genome-wide screening and characterization of long non-coding RNAs involved in flowering development of trifoliate orange (*Poncirus trifoliata L. Raf.*)[J]. Sci Rep, 7:43226.

XIONG Y Y , Ma J , He Y H , et al. ， 2018. High-throughput sequencing analysis revealed the regulation patterns of small RNAs on the development of *Ananas comosus* var. *bracteatus* leaves[J]. Sci Rep, 8(1):1947.

XIONG YY, 2019. MicroRNAs identification and screening and functional verification of key microRNAs involved in the albino of *Ananas comosus* var. *Bracteatus*[D]. Ya'an: Sichuan Agricultural University: 1-95.[熊颖媛，2019. 红苞凤梨 microRNA 的鉴定及白化关键 microRNA 的筛选与功能验证[D]. 雅安：四川农业大学: 1-95.]

XUE Y , MA J , HE Y , et al. ， 2019. Comparative transcriptomic and proteomic analyses of the green and white parts of chimeric leaves in *Ananas comosus* var. *bracteatus*[J]. PeerJ, 7:e7261.

XU WN, HUANG RM, LIU YY, et al. ， 2018. Genome sequencing and assembly strategy analyses of *Flammulina filiformis*[J]. Mycosystema, 37(12):1578-1585. [徐伟南， 黄蓉梅，刘媛媛，等，2018. 金针菇基因组测序与组装策略分析[J]. 菌物学报，37(12):1578-1585.]

YU CL, LUO L, LIAO Q, 2015. Annotation and Functional Prediction of lncRNAs[J]. Chin J BiochemMol Biol, (3): 239-243. [余铖亮，骆亮，廖奇，2015. lncRNAs 功能注释和预测[J]. 中国生物化学与分子生物学报, (3): 239-243.]

ZHANG Y, TAO Y, LIAO Q, 2018. Long noncoding RNA: A crosslink in biological regulatory network[J]. Brief Bioinform, 19(5): 930-945.